数理解析研究所講究録
611巻 1987年 69-89

ある部分観測可能なマルコフ過程における最適停止問題について

大阪府立大学総合科学部
中井　達 (Tōru Nakai)

§1. Introduction

逐次に出現する確率変数の列に対して、それらの実現値を観測することが出来、さらにその実現値を得ることが出来る場合の最適停止問題について考える。ここで出現する確率変数列の各々の確率変数は discrete time Markov process { Zn; n=0,1,2,.... } に従って変化するこのマルコフ過程の状態に依存するものと考え、その状態を decision-maker は直接に観測することは出来ず、逐次に出現する確率変数の実現値を観測することにより、そのマルコフ過程の状態に関する情報を得るものと考える。その状態に関する情報を用いた学習の方法は、ここではベイズの方法によって学習する場合について考える。しかし、よく知られているようにベイズの方法に従って学習を行う場合においては、事後確率に関して

色々と複雑な問題が起きて来る。そのためにここではいくつかの基本的な仮定を設け、その仮定のもとで事後確率がどの様な動きをするかを観察し、その結果をもとにして、先に述べた部分観測が可能なマルコフ過程における最適停止問題について考えてみる。

§2. 基本的な仮定

Discrete time Markov process を{ Z_n ; n = 0,1,2,....}とし、そのマルコフ過程の状態空間は Borel set S の値を取るものとする。ここでは S は R^d の Borel subset でコンパクトかつ連結なものと仮定し、また S においては順序が定義されており、その順序は全順序であるとする。しかしより一般の metric space においても同様にコンパクト性、連結性及び全順序を仮定すれば同様の議論が出来る。また R^1 の部分空間を考える場合はコンパクト性は必要とはならない。つぎに A を S の任意の Borel subset とし、S に含まれる任意の s にたいして $p_n(s;A)$ をこのマルコフ過程の Markovtransition function とする。また同様に stationary Markov transition function は $p(s;A)$ で表すものとする。以下の議論においては記号の煩雑さを省くため stationary Markov proces の場合について扱うが、そうでな

い場合についても同様に扱うことが出来る。

まず S の二つの Borel set B_1, B_2 にたいして順序を次のように定義する。（但し集合 S には、上に述べたように全順序が定義されている）即ち

"$B_1 < B_2$ if and only if $b_1 < b_2$ for any b_1 of B_1 & any b_2 of B_2"

とする。この時、次の仮定を設ける。

仮定 1.

$$p(s_1;B_1)p(s_2;B_2) < p(s_2;B_1)p(s_1;B_2) \quad \text{for } s_1 < s_2,\ s_1,\ s_2 \in S.$$

この仮定は、よく知られているマルコフ過程における order 2 の totally positivity と同様の仮定である。次に、S に含まれる各々の値 s に対して実数値の確率変数 X_s を考え $F_s(x)$ を X_s の probability distribution function とし、$F_s(x)$ は absolutely continuous であると仮定する。このことは仮定しなくても以下の議論は成立するが、簡単のためここではこの仮定を設け、また $f_s(x)$ は X_s の density function とする。ここで $f_s(x)$ は s に関して一様有界でかつ連続であるものと考える。この時、更に次の三つの仮定を

考える。

仮定 2.

　Sに含まれる任意の s に対して $E[X_s] < \infty$ が成り立つ。

仮定 3.

　Sに含まれる任意の二つの値 s, t に対して $s > t$ であるならば不等式

$$f_t(x)f_s(y) \leqq f_s(x)f_t(y)$$

が全ての x 及び y ($x < y$) に対して成り立つ。

　この仮定はマルコフ過程において考えた先の仮定 1 と同様の order 2 の totally positivity を density function に対して、仮定したものである。

仮定 4.

　Sに含まれる全ての s, t に対して $t < s$ であるならば $x_{ts} = \sup\{ x \mid f_t(x) \leqq f_s(x) \}$ となる値 x_{ts} が存在して次の事柄が成り立つ。即ち

$x > x_{ts}$ ならば $f_t(x) > f_s(x) > f_{s'}(x)$ for any $s' > s$,

$x \leqq x_{ts}$ ならば $f_s(x) \geqq f_t(x) \geqq f_{t'}(x)$ for

any $t' < t$.

ここでは X_s は実数値の確率変数として扱っているが、一般の位相空間 T（$\mathcal{B}$(T) は加算族で生成されるものとする）に於て X_s が T 値の確率変数であると考えたとき、仮定 3 及び 4 は同様の形で表すことが出来る。（但し T には適当な順序が定義されているものとする）

次に状態空間 S 上の measure を ρ とする。この時マルコフ過程の状態に関する prior information が S 上の測度 ρ であるとき確率変数 X_s（$Z = s$）の実現値 x を観測したときの posterior information $\bar{T}_\rho(x)$ はベイズの定理により次のような S 上の measure として表すことが出来る。（但しここでは prior information は前の期の後で推移が起こった以後において得られているものであるとする）

B を S の Borel subset とするとき、ベイズの定理に従って

$$T_\rho(B \mid x) = \frac{\int_B f_s(x)\,\rho(ds)}{\int_S f_s(x)\,\rho(ds)} \tag{1}$$

及び

$$\bar{T}_{\mathfrak{p}}(B \mid x) = \int_S p(s;B)\, T_{\mathfrak{p}}(ds) \tag{2}$$

と表すことが出来る。

§3 A Partial order

次に $(S, \mathfrak{B}(S))$ 上の probability measure $\mathfrak{p}$ の全体に対して次のような順序を導入することを考える。

定義

$\mathfrak{p}$, $\mathfrak{q}$ を共に S 上の probability measure であるとする。この時 $\mathfrak{p} \geqq \mathfrak{q}$ であるとは次の性質が成り立つことである。即ち

B_1, B_2 が $\mathfrak{B}(S)$ に含まれる二つの Borel subset であって $B_1 < B_2$ であれば不等式

$$\mathfrak{p}(B_1)\mathfrak{q}(B_2) \geqq \mathfrak{q}(B_1)\mathfrak{p}(B_2)$$

が成り立つ。

このとき S 上の probability measure $\mathfrak{p}$ の間には partial order が定義されることになり、以下のような性質を得ることが出来る。

Lemma 1.

$\wp$, q を S 上の probability measure であるとする。もし $\wp \geq q$ であるならば次の性質が成立する。即ち

S の部分集合 S_1、S_2 で $S_1 \cap S_2 = \emptyset$, $S_1 \cup S_2 = S$ となるものが存在して次の事柄が成り立つ。（S_1, S_2 は共に connected である）

B が $\mathcal{B}(S)$の要素でありもし

1) $B \subset S_1$ であれば $\wp(B) > q(B)$が成立ち、一方

2) $B \subset S_2$ であるならば $\wp(B) < q(B)$ が成り立つ。

（略証）まず 1) を満足する S の部分集合を S_1 とし、また 2) を満足する S の部分集合を S_2 とする。

最初にこのような二つの集合が存在することを示す。いま任意の S に含まれる Borel set B に対して B^U および B^L を次のように定義する。即ち、

$$B^U = \{ s \mid s \in S,\ s > b \text{ for any } b \text{ in } B \},$$

$$B^L = \{ s \mid s \in S,\ s < b \text{ for any } b \text{ in } B \}$$

とする。このとき次のうち少なくとも一方が成り立つ。 a) S に含まれる Borel set B で $B^U \neq \emptyset$ かつ $\wp(B) - q(B)$

> 0 である。b) S に含まれる Borel subset B で $B^{L} \neq \emptyset$ かつ $p(B) - q(B) < 0$ である。S 上の probability measure の間における partial order の定義に従えば、$q \leq p$ かつ $B_1 < B_2$ であれば

$$q(B_1)(p(B_2) - p(B_1)) < p(B_1)(q(B_2) - q(B_1))$$

がなりたつ。従ってもし a) が成り立てば S_1 が存在することが分かり、一方 b) が成り立てば S_2 の存在が示される。次に S_1 または S_2 の存在することが分かれば、それぞれに対応して S_2 または S_1 が存在することは p および q が probability measure であることに注意すれば比較的簡単に示すことが出来る。

次に S_1 および S_2 の存在に付いては示すことが出来たが、次に状態空間 S が二つの S_1 および S_2 に分けられるかどうか、即ち $S = S_1 \cup S_2$ であるかどうかを考える。いま、$(S_1 \cup S_2)^c$ が空集合でないとき、$(S_1 \cup S_2)^c$ に含まれる空でない任意の部分集合 B を考える。このとき $p(B) - q(B) > 0$ または $p(B) - q(B) < 0$ である。ここで最初の場合を考える。このとき B に含まれる Borel subset の列 $\{A_n\}$ および $\{C_n\}$ で次のような性質を持つものが存在することが示される。即ち、

$$A_1 \subset A_2 \subset \cdots\cdots \subset A_n \subset \cdots\cdots,$$

$$\wp(A_n) - \mathscr{q}(A_n) < 0 \text{ for any } n > 0,$$

かつ

$$C_1 \supset C_2 \supset \cdots\cdots \supset C_n \supset \cdots\cdots,$$

$$\wp(C_n) - \mathscr{q}(C_n) > 0 \text{ for any } n > 0,$$

となる。更に $A = \bigcup A_n$ および $C = \bigcap C_n$ とおけば $\wp(A) - \mathscr{q}(A) < 0$ かつ $\wp(C) - \mathscr{q}(C) > 0$ であり更に C に含まれる任意の D に対して $\wp(D) - \mathscr{q}(D) > 0$ となることを示すことが出来る。従って C は S_1 に含まれなければならず、仮定に反することが言える。

この lemma は partially observable Markov chain における Nakai[4] において得られた同様の性質（ Lemma 3.2)の一般化となっている。

次に S 上の measure の上の関数 $u(\wp)$ を考える。（例えば $u(\wp) = \int_S g(s)\wp(ds)$ の様に integrable function g(s) を積分した値といったような関数を考える ）この時 $u(\wp)$ が $\wp$ に関して増加関数であるとは上の定義に於て述べ

た意味で $\mathfrak{p} \geqq \mathfrak{q}$ であるならば $u(\mathfrak{p}) \geqq u(\mathfrak{q})$ であることを言う。このとき次の Lemma が成り立つ。

Lemma 2.

関数 $a(s)$ が s に関して減少関数であり、また関数 $v(\mathfrak{p}, s)$ が s に関して減少かつ $\mathfrak{p}$ に関して増加関数であるならば

$$u(\mathfrak{p}) = \int_S a(S)v(\mathfrak{p}, s)\mathfrak{p}(ds)$$

は $\mathfrak{p}$ に関して増加する関数である。即ち $\mathfrak{p} \geqq \mathfrak{q}$ であれば $u(\mathfrak{p}) \geqq u(\mathfrak{q})$ が成り立つ。

この lemma は lemma 1 を用いて Nakai [4] で扱ったと同様の方法により示すことが出来る。

次に S に関する prior information が $\mathfrak{p}$ であるとき実現値 x を観測したという条件のもとでの posterior information $T_{\mathfrak{p}}(x)$ および $\overline{T}_{\mathfrak{p}}(x)$ に関して次の三つの性質を得ることが出来る。

Lemma 3.

もし $x \geqq y$ であるならば、任意の S 上の

probability measure p に対して

$$\bar{T}_p(x) \geqq \bar{T}_p(y) \qquad (3)$$

が成り立つ。

このlemmaは(1)式、及び仮定3を用いて示すことが出来る。即ち、任意の二つのBorel subset B_1, B_2 で $B_1 \leqq B_2$ であるものに対して不等式

$$\bar{T}_p(B_1 \mid x)\bar{T}_p(B_2 \mid y) \leqq \bar{T}_p(B_2 \mid x)\bar{T}_p(B_1 \mid y)$$

が成り立つことを示せば良いが、これは仮定3により容易に示すことが出来る。

Lemma 4.

もしS上の二つのprobabilty measure p および q に対して $p \geqq q$ であるならば

$$T_p(x) \geqq T_q(x) \qquad (4)$$

が R^1 に含まれる全ての x に対して成り立つ。

このlemmaは(1)式とS上のprobability measureのpartial orderに関する定義により示すことが出来る。

即ち、

$$T_p(B_1 \mid x)T_q(B_2 \mid x) - T_q(B_1 \mid x)T_p(B_2 \mid x) \geq 0$$

の分母を払ったものが正であることが示される。ただし B1 ≦ B2 であるものとする。

Lemma 5.

もし S 上の二つの probability measure p, q に対して $p \geq q$ であるならば

$$\bar{T}_p(x) \geq \bar{T}_q(x) \tag{5}$$

が R^1 に含まれる全ての x に対して成り立つ。

Lemma 5 は先に示した Lemma 4 を用いて Lemma 4 と同様にして、S 上の probability measure に於ける partial order の定義に従って求めることが出来る。またここにおいて S 上に total order が定義されていることが、この lemma を示すときに必要になってくる。

次の性質は、最適停止問題を考える場合に重要となる性質であるが、この節で扱っている S 上の probability measure における partial orderと関連してここで述べることにする。

Lemma 6.

$\{f_s(x)\}$ を前の節で仮定した仮定 4 と同様の性質を持つような density function の列とし、さらに S 上の二つの probability measure $\wp$, $\mathfrak{q}$ に対して $\wp \geqq \mathfrak{q}$ であるとする。このとき次の性質が成り立つ。

即ち、

$$g(x) = \int_S f_s(x)(\wp(ds) - \mathfrak{q}(ds)) \tag{6}$$

とおくとき、任意の x に関する増加関数 $h(x)$ に対して

$$\int_{R^1} h(x)g(x)dx \geqq 0$$

が成り立つ。

（略証）まず二つの S 上の probability measure $\wp$ 及び $\mathfrak{q}$ に対して次のような二つの S 上の probability measure の列 $\{\wp_n\}$, $\{\mathfrak{q}_n\}$ が存在することが示される。$\wp_n$ 及び $\mathfrak{q}_n$ はともに discete measure であってかつ $\wp_n \geqq \mathfrak{q}_n$ であり、更に Doob [1], Feller [2] で考えられた意味においてそれぞれ $\wp$ 及び $\mathfrak{q}$ に収束するものである。

つぎに $g_n(x)$ を

$$g_n(x) = \int_S f_s(x)\,(\wp(ds) - \mathcal{q}(ds))$$

により定義をすれば

$$\lim_{n\to\infty} g_n(x) = g(x)$$

であることが示される。さらに Nakai [4] において得られた partially observable Markov chain に関する結果（ Lemma 3.4 ）を用いて

$$\int_{R^1} h(x) g_n(x) dx \geqq 0$$

が x に関して増加する任意の関数 $h(x)$ に対して成り立つことがわかる。従って、$\wp_n$ 及び $\mathcal{q}_n$ がそれぞれ $\wp$ および $\mathcal{q}$ に収束することにより

$$\int_{R^1} h(x) g(x) dx = \lim_{n\to\infty} \int_{R^1} h(x) g_n(x) dx \geqq 0$$

であることが示され、この lemma が得られる。

§4. 最適停止問題

ここでは次のような問題について考えることにする。今 $(\wp, N)$ を状態に関する事前の情報が $\wp$ であり、残りの決定期間の数が N であるときのこの部分観測が可能なマルコフ

過程における最適停止問題とする。このとき、($\wp$, N)をこの最適停止問題の状態と考え、この問題を $P_N(\wp)$ で表す。またこの最適停止問題が状態($\wp$, N)である時、最適政策に従ったという条件のもとでの期待値を $V_N(\wp)$ で表し、この問題の値という。

このとき、最適停止問題 $P_N(\wp)$ は次のようにして行われると考える。まずこの最適停止問題の状態が($\wp$, N)であるとき decision-maker は確率変数 X の実現値 x を観測する。ここでこの確率変数は第1節で考えたように、あるマルコフ過程に従って変化する状態に依存するが、その状態を直接には知ることは出来ず、あらかじめ与えられた事前知識によってのみその状態に関する情報を得ることが出来る。このとき decision-maker はこの確率変数の実現値 x を取って stop するか、それとも更に次の stage へ進み次の確率変数を観測するかを決定する。もし stop することを決定したならば immediate reward $u(\wp, x)$ を得るものとする。(但し $u(\wp, x)$ は $\wp$ に関しては増加する関数であり、また x に関しても増加する関数であるものとする)もし stop しないのならばこの最適停止問題は、次の stage へと進み、残りの決定期間の数は N - 1 となり、マルコフ過程の状態に関する information はベイズの方法に従い $\overline{T}\wp(x)$ と improve され、

この最適停止問題の状態は（N-1, $\bar{T}_{\wp}(x)$）となる。ここで immediate reward $u(\wp, x)$ は $\wp$ に関して増加する関数であることを仮定したが、この仮定には同じ値であっても状態に関する情報が異なればその価値は当然変化することを考慮したものである。例えば、具体的に言えばマルコフ過程の状態を経済状況を示す指標といったものと考えるとき、同じ値であっても、景気の良いときと悪いときとではその価値は異なっていると考えられることである。

従って Ross [6] に於て述べられたと同様の方法に従い先に述べた部分観測が可能なマルコフ過程における最適停止問題を Markov decision process として定式化すれば次のような最適方程式を得ることが出来る。

$$V_N(\wp) = \int_{R^1} V_N(\wp \mid x)\, dF_{\wp}(x) \tag{7}$$

$$V_N(\wp \mid x) = \max \{ u(\wp, x),\ V_{N-1}(\bar{T}_{\wp}(x)) \} \tag{8}$$

$$V_1(\wp \mid x) = u(\wp, x) \tag{9}$$

である。但し $F_{\wp}(x)$ は weighted distribution function

$$F_{\wp}(x) = \int_S f_s(x)\, \wp(ds)$$

であり、また $f_s(x)$ に対する仮定によりこの関数は well

defined である。従って、(7) 式はまず出現した確率変数 X の実現値 x を観測し、その値をもとにして決定を行うことを表す。Decision-maker の決定は (8) 式において示されているように、実現値 x をとって stop し、immediate reward $u(\beta, x)$ を得るか、それとも次の値を観測するかを決定する。もしも次の値を観測するならば、マルコフ過程の状態は与えられた Markov transition function に従って次の状態へと移り、そのときの posterior information は $\bar{T}\beta(x)$ となり、その時の期待値は $V_{N-1}(\bar{T}\beta(x))$ となる。

このとき Nakai [3], [4] において得られたと同様に、前節で示された性質を用いて、数学的帰納法により次にあげた性質を得ることが出来る。これらの proposition は immediate reward $u(\beta, x)$ が β に関して増加する関数であること、及び posterior probability distribution $\bar{T}\beta(x)$ が β に関して増加するという Lemma 4 の結果を用いて示すことが出来るものである。

Proposition 1.

最適停止問題 $P_N(\beta)$ の値 $V_N(\beta)$ は β に関して増加する関数である。

Proposition 2.

全ての S 上の probability measure $\wp$ に対して $V_N(\wp) \geqq V_{N-1}(\wp)$ である。

Proposition 1 はマルコフ過程の状態に関する情報が第3節で定義した partial order の意味に於て大きければ大きいほど、ここで考えた最適停止問題の値は大きいことを示している。従ってこの意味においてここで仮定した条件は不自然ではないものと考えられる。Proposition 2 はこの最適停止問題に於て、残りの決定期間の数が多ければ多いほど、得られる値は大きいことを示している。

次に stopping region $S_N(\wp)$ を以下のように定める。即ち

$$S_N(\wp) = \{ x \mid u(\wp, x) \geqq V_{N-1}(T_\wp(x)) \}. \qquad (10)$$

この stopping region $S_N(\wp)$ は最適停止問題の状態が $(\wp, N)$ であるとき、出現した確率変数の実現値 x が $S_N(\wp)$ に含まれるならば decision-maker は stop してこの値を取る。一方、反対に x が $S_N(\wp)$ に含まれなければ decision-maker はこの値を見逃して、次の stage へと進み次の確率変数の値を observe する。従ってこの stopping region $S_N(\wp)$ は、ここで考えた最適停止問題の最適政策を特徴付けるものであり、最適政策はこれにより求めることが出来る。

このとき Propositions 1 及び 2 において得られた性質をもとにして次の性質が成り立つ。

Proposition 3.

任意の $N(\geq 1)$ に対して

$S_N(\wp) \subset S_{N-1}(\wp)$ である。

この proposition は $V_{N-1}(\bar{T}_\wp(x))$ が N に関して増加する関数であるという Proposition 2 の結果と stopping region $S_N(\wp)$ の定義から簡単に求めることが出来る。またこの proposition により、残りの決定期間の数、即ち stage 数 N が大きければ大きいほど、この最適停止問題の optimal policy はより厳しい policy に従うことが分かる。即ち、stopping region $S_N(\wp)$ は N が大きければ大きいほどより小さくなることを示している。

Proposition 4.

immediate reward $u(\wp, x)$ が $\wp$ に依存しない場合、

即ち $u(\beta, x) = u(x)$ である場合には、もし $\beta \geq \eta$ であるならば

$S_N(\beta) \subset S_N(\eta)$ である。

この proposition は $V_{N-1}(\beta)$ が β に関して第3節において定義した partial order の意味において増加する関数であることと、immediate reward $u(\beta, x)$ が β に依存しないというこの proposition の仮定から容易に求めることが出来る。従って、この proposition からマルコフ過程の state に関する事前知識が先に述べた意味において大きければ大きいほど、この最適停止問題の optimal policy を決定する stopping region $S_N(\beta)$ はより小さくなることが示される。

ここでは、部分観測が可能なマルコフ過程における最適停止問題を扱ったが、更に第2節において仮定したと同じ状態のもとで、job search の問題、最適取り替え問題などの逐次決定問題に関しても同様の部分観測が可能なマルコフ過程において、ベイズの方法に従って学習を行う場合に同様の性質を得ることが出来る。

References

1. J.L.Doob, " Stochastic Processes ", 1953, John Wily & Sons,New York.

2. W. Feller, " An Introduction to Probability Theory and Applications I " 1950, John Wily & Sons, Newe York.

3. T. Nakai, " Optimal Stopping Problem in a Finite State Partially Observable Markov Chain ", Journal of Information & Optimization Sciences, vol. 4, 1983, pp. 159-176.

4. T. Nakai, " The Problem of Optimal Stopping in a Partially Observable Markov Chain ", Journal of Optimization Theory and Applications, vol. 45, 1985, pp. 425-442.

5. T. Nakai, " A Sequential Stochastic Assignment Problem in a Partially Observable Markov Chain ", Mathematics of Operations Research, vol. 11, 1986, pp. 230-240.

6. S.M. Ross " Applied Probability with Optimization Applications ", 1970, Holden Day, New York.